ISBN 978-989-9121-70-6
30th PRAGUE International Conference on Electrical and Electronics Engineering (PE3-26)
July 27-29, 2026 Prague (Czech Republic)

Lightweight CNN-Based Acoustic Beehive Health Monitoring System Using Raspberry Pi

Li-Hao Chen, Ping Chia Wu, and Chau-Chung Song*

*Chau-Chung Song is with the department of Aeronautical Engineering and Smart Machine and Intelligent Manufacturing Research Center, National Formosa University, Taiwan

Li-Hao Chen is with the graduate institute of Aviation and Electronic Technology, National Formosa University, Taiwan

Ping Chia Wu is with the graduate institute of Aviation and Electronic Technology, National Formosa University, Taiwan

Abstract: *This study presents an integrated, non-invasive beehive monitoring framework leveraging a three-layer architecture comprising sensing, Raspberry Pi edge computing, and cloud visualization layers.*

Utilizing acoustic emission patterns as a diagnostic proxy for queen status, the study evaluates the efficacy of four lightweight convolutional neural networks (MobileNetV1–V4) across three distinct time-frequency representations. Experimental evaluations using the Kaggle Beehive Sounds Dataset demonstrate that the MobileNetV4 architecture coupled with Gammatone Cochleagram features yields superior baseline performance. Following systematic Optuna-based hyperparameter optimization, this configuration achieves a 97.34% mean accuracy under five-fold cross-validation.

The empirical findings validate that optimized lightweight CNNs successfully reconcile the trade-off between high classification accuracy and resource-constrained embedded deployment within precision apiculture applications.

Keywords: Acoustic Beehive Monitoring, Edge Computing, Gammatone Cochleagram, Lightweight CNN, Precision Apiculture

1. Introduction

Continuous beehive monitoring has attracted increasing attention as a non-invasive approach for assessing colony status and supporting precision apiculture. Conventional hive inspection relies heavily on manual opening and visual observation; therefore, it is labor-intensive, intermittent, and may disrupt colony activity. Recent studies have shown that acoustic patterns can provide useful information for bee colony monitoring and can be classified using machine learning models on constrained devices [1], [2]. These findings motivate the development of continuous monitoring approaches capable of operating directly in field environments.

To address this need, this paper presents an integrated beehive health monitoring system that combines embedded sensing, IoT-based communication, and lightweight deep learning inference on Raspberry Pi. The proposed system continuously collects hive-related acoustic and environmental information, transmits structured results to a cloud database, and provides remote visualization through a web-based dashboard. In this study, queen-status-related acoustic categories are used as a practical proxy indicator of colony condition, allowing the system to infer biologically meaningful hive states without physically disturbing the bees. The main contributions of this work are threefold: (1) an integrated edge-to-cloud beehive monitoring system, (2) a unified comparison of three acoustic feature representations and four MobileNet models, and (3) Optuna-based hyperparameter optimization for MobileNetV4 with Gammatone Cochleagram.

2. Methodology

2.1. System Architecture

The proposed system adopts a three-layer IoT architecture as illustrated in Fig. 1. The Sensing Layer consists of four sensors connected to the Raspberry Pi 4B: a USB microphone for acoustic acquisition, an SGP30 gas sensor for volatile organic compound (VOC) monitoring, an AHT20 temperature and humidity sensor connected through I²C, and an HX711-based weight sensing module. Together, these devices provide complementary acoustic and environmental information for hive condition assessment.

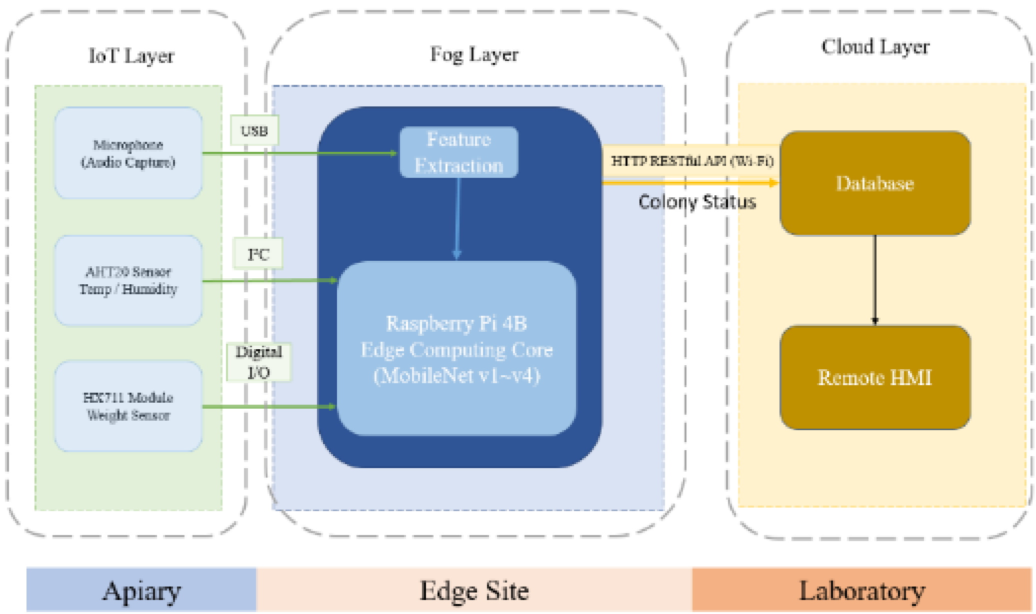


Fig. 1. System architecture of the beehive health monitoring system based on sensing, edge computing, and cloud layers

2.2.Raspberry Pi System Configuration

The Raspberry Pi 4B (shown in Fig. 2) runs a Linux-based operating system configured for audio acquisition and on-device inference. The USB microphone records beehive audio at a sampling rate of 16,000 Hz, and each recording is segmented into 5-second clips for downstream feature generation and classification. In parallel, the AHT20 and SGP30 sensors monitor temperature, relative humidity, and gas-related indicators, while the HX711 module measures hive weight as an indirect indicator of nectar/honey accumulation and colony activity. This combination enables the platform to support both acoustic intelligence and contextual environmental monitoring.

https://doi.org/10.17758/HEAIG19.ER0726701

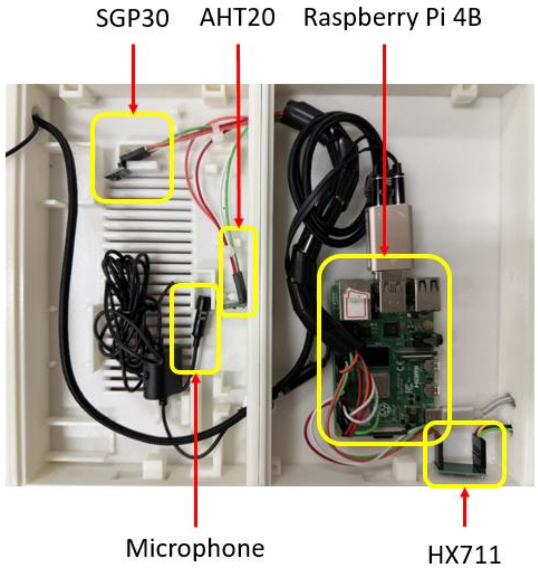


Fig. 2. Hardware components and sensor placement of the monitoring system

HTTP RESTful API Communication Protocol and Data Flow

The system employs HTTP RESTful API communication over Wi-Fi as the primary data exchange mechanism, providing a stateless and platform-independent interface between the edge node and the cloud server. Sensor readings and inference outputs are serialized as JSON payloads and uploaded to the cloud server through HTTP POST requests, while configuration updates and status queries are retrieved through HTTP GET requests. This design keeps the edge node lightweight and simplifies integration with database and web services without introducing additional broker infrastructure.

2.4. Cloud Integration

The Cloud Layer includes a relational database for long-term storage of sensor readings and classification results, as well as a web-based Human–Machine Interface (HMI) for remote visualization. The dashboard presents current hive status, historical trends, and threshold-based alert information, allowing users to inspect colony conditions without requiring on-site intervention.

2.5. Audio Feature Extraction and Model Training

The Kaggle Beehive Sounds Dataset is used to evaluate three audio feature extraction methods. Spectrogram provides a standard time–frequency representation based on the Short-Time Fourier Transform (STFT). Log Mel Spectrogram maps frequency components onto the perceptually motivated Mel scale. Gammatone Cochleagram is a biologically inspired representation generated by a gammatone filter bank with 64 channels and a hop time of 0.010 s, producing approximately 500 time frames for each 5-second audio clip. This representation is suitable for emphasizing harmonic structures relevant to bee wing-beat acoustics [7], [8].

For a fair comparison, all baseline models are trained using the same optimization pipeline.MobileNetV1–V4 are initialized from ImageNet-pretrained weights via the timm library with single-channel input support (in_chans=1), and then fine-tuned using a fixed baseline protocol including batch size 64 learning rate=1×10-5,the AdamW optimizer(weight_decay=1×10-3), ReduceLROnPlateau scheduler (factor = 0.5, patience = 8), dropout 0.5, drop path rate 0.2, label smoothing 0.1, early stopping patience 20, and a maximum of 250 epochs [3]–[6]. The dataset is divided into training and validation subsets using an 80%/20% split with group-wise separation by recording ID to reduce data leakage. Gaussian noise and frequency masking are applied only to the training subset as augmentation strategies. The final task is four-class queen status classification (classes 0–3), which serves here as an acoustically observable proxy for colony condition.

Although the present study focuses on acoustic classification performance, the comparison is designed with edge deployment in mind. All evaluated backbones belong to the lightweight MobileNet family, making them suitable candidates for future real-time inference on resource-constrained embedded platforms such as Raspberry Pi. Accordingly, the analysis emphasizes both classification accuracy and the practicality of selecting compact architectures for precision apiculture applications.

2.6. Hyperparameter Optimization

MobileNetV4 with Gammatone Cochleagram is selected as the optimization target for two reasons: its biologically inspired frequency representation aligns well with bee wing-beat acoustics, and its highest baseline accuracy of 96.28% suggests further potential for systematic tuning. Hyperparameter optimization is conducted using the Optuna framework. The Tree-structured Parzen Estimator (TPE) sampler is adopted to search the hyperparameter space, while a median pruning strategy is used to terminate underperforming trials and reduce unnecessary computational cost [9]. The optimization objective is defined as maximizing the mean validation accuracy obtained from five-fold cross-validation. To ensure a fair before-and-after comparison, the baseline MobileNetV4 with Gammatone Cochleagram configuration is also re-evaluated using the same five-fold cross-validation protocol. Notably, the five-fold cross-validation baseline accuracy of 96.28% coincides with the 80%/20% hold-out result reported in Table I, although the two evaluations are conducted under different protocols.

The search space includes learning rate, weight decay, drop rate, drop path rate, and scheduler type. In addition, per-fold balanced class weights are applied during training to reduce the influence of class imbalance. To enable efficient search within a limited computational budget, each Optuna trial is trained with a reduced schedule of a maximum of 60 epochs and an early stopping patience of 12, in contrast to the 250-epoch baseline protocol. The best configuration is obtained at Trial 19,with a learning rate 1.700×10-4 . The model is trained using the AdamW optimizer with a batch size of 64, label smoothing of 0.1, a maximum of 60 epochs, and early stopping patience of 12. The remaining optimal hyperparameters of Trial 19(weight_decay= 6.361×10^{-4} ,drop rate= 0.2511, drop path rate= 0.2007 and scheduler type is cosine warmup)

3. Results and Discussion

Table I summarizes the baseline validation performance of all twelve model-feature combinations in terms of macro-precision, macro-recall, macro-F1 score, and accuracy. Among them, MobileNetV4 with Gammatone Cochleagram achieves the best baseline performance, with 96.28% accuracy, 96.32% macro-precision, 96.28% macro-recall, and 95.49% macro-F1 score. MobileNetV4 with Spectrogram achieves the second-highest Spectrogram-based accuracy of 94.69%, indicating that the standard STFT-based representation also provides competitive performance for bee acoustic classification.

TABLE I. PERFORMANCE COMPARISON OF MOBILENETV1–V4 ACROSS THREE FEATURE EXTRACTION METHODS

Model	Prec.	Rec.	Macro-F1	Acc.
MNv1 + Spec	0.9445	0.9445	0.9316	0.9445
MNv2 + Spec	0.9429	0.9424	0.9290	0.9424
MNv3 + Spec	0.9413	0.9403	0.9266	0.9403
MNv4 + Spec	0.9469	0.9469	0.9353	0.9469
MNv1 + Mel	0.9406	0.9399	0.9263	0.9399
MNv2 + Mel	0.9472	0.9468	0.9349	0.9468
MNv3 + Mel	0.9521	0.9519	0.9416	0.9519
MNv4 + Mel	0.9566	0.9565	0.9468	0.9565
MNv1 + Gamma	0.9559	0.9556	0.9456	0.9556
MNv2 + Gamma	0.9586	0.9583	0.9494	0.9583
MNv3 + Gamma	0.9609	0.9606	0.9529	0.9606
MNv4 + Gamma	0.9632	0.9628	0.9549	0.9628

Although MobileNetV4 with Gammatone Cochleagram achieves the best baseline performance among the twelve model-feature combinations, it is further selected for hyperparameter optimization because of its biologically inspired acoustic representation and strong baseline accuracy. As shown in Fig.3, the baseline MobileNetV4 with Gammatone Cochleagram achieves 96.28% accuracy, whereas the optimized configuration achieves 97.34% accuracy under five-fold cross-validation. This corresponds to an absolute improvement of 1.06 percentage points.

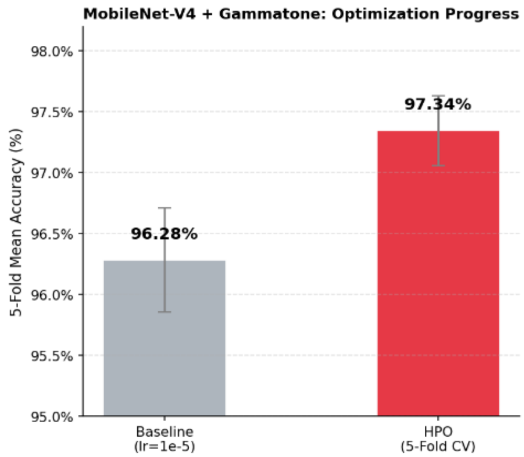


Fig. 3. MNv4 + Gammatone: Baseline and Optimized Performance Comparison

Table II further compares the macro-level performance before and after hyperparameter optimization. The optimized configuration improves precision from 0.9632 to 0.9690, recall from 0.9628 to 0.9679, and macro-F1 score from 0.9549 to 0.9683.

TABLE II. PERFORMANCE IMPROVEMENT OF MOBILENETV4 WITH GAMMATONE COCHLEAGRAM AFTER HYPERPARAMETER OPTIMIZATION

Model	Prec.	Rec.	Macro-F1	Acc.
Baseline	0.9632	0.9628	0.9549	0.9628
HPO	0.9690	0.9679	0.9683	0.9734

The improvement after hyperparameter optimization indicates that the performance gain is not limited to overall accuracy but is also reflected in more balanced multi-class classification behavior. In particular, the increase in macro-F1 score suggests that the optimized MobileNetV4 with Gammatone Cochleagram provides more stable recognition across different queen-status classes. This result further supports the suitability of Gammatone Cochleagram for bee acoustic analysis, as its filter-bank structure can preserve discriminative harmonic patterns related to wing-beat acoustic differences.

Several trends can be observed from the experimental results. First, as shown in Table I, Gammatone Cochleagram produces the strongest baseline performance across all MobileNet variants, with MobileNetV4 achieving the highest accuracy of 96.28%. This suggests that the biologically inspired frequency representation preserves discriminative harmonic structures that are relevant to bee wing-beat acoustics. Second, Spectrogram yields competitive but lower performance than Gammatone Cochleagram under the baseline setting, with MobileNetV4 achieving 94.69% accuracy. Although the STFT-based representation is computationally simpler, it appears to capture slightly less discriminative information for this task. Third, Log Mel Spectrogram shows intermediate performance among the three feature types, with MobileNetV4 achieving the best Mel result at 95.65% accuracy. Overall, all MobileNet variants with Mel features fall between the Spectrogram and Gammatone results, suggesting that Mel-scale frequency compression retains useful acoustic information but does not match the discriminative capability of the gammatone filter bank for queen-status classification in this dataset.

Across all experiments, MobileNetV4 shows the most promising overall behavior. It achieves the best baseline result with Gammatone Cochleagram and further improves to the highest final accuracy after hyperparameter optimization. Since the MobileNet family is designed for efficient CNN-based inference, MobileNetV4 is considered the most suitable backbone for future embedded deployment in the proposed Raspberry Pi-based beehive monitoring system [3]–[6].

4. Conclusion

This study presents a lightweight acoustic beehive health monitoring framework integrating multi-sensor data acquisition, Raspberry Pi edge computing, cloud-based visualization, and CNN-based queen-status classification. Among the evaluated model-feature combinations, MobileNetV4 with Gammatone Cochleagram achieved the best baseline performance, with an accuracy of 96.28%. After Optuna-based hyperparameter optimization, the same configuration further improved to 97.34% mean accuracy under five-fold cross-validation, demonstrating the effectiveness of combining biologically inspired acoustic representation with lightweight CNN architectures.

These results indicate that acoustic feature-based lightweight deep learning is a feasible approach for non-invasive beehive monitoring. Future work will focus on real-device inference latency, power consumption evaluation on Raspberry Pi, long-term field validation, and extension toward multi-colony monitoring with additional environmental sensing modalities.

5. References

[1] Robles-Guerrero, A., Gómez-Jiménez, S., Saucedo-Anaya, T., López-Betancur, D., Navarro-Solís, D., & Guerrero-Méndez, C. (2024). Convolutional Neural Networks for Real Time Classification of Beehive Acoustic Patterns on Constrained Devices. Sensors, 24(19), 6384.
https://doi.org/10.3390/s24196384

[2] Robles-Guerrero, A., Saucedo-Anaya, T., Guerrero-Mendez, C. A., Gómez-Jiménez, S., & Navarro-Solís, D. J. (2023). Comparative Study of Machine Learning Models for Bee Colony Acoustic Pattern Classification on Low Computational Resources. Sensors, 23(1), 460.
https://doi.org/10.3390/s23010460

[3] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv, abs/1704.04861.

[4] Sandler, M., Howard, A.G., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4510-4520.
https://doi.org/10.1109/CVPR.2018.00474

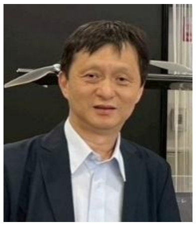
[5] Howard, A.G., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., & Adam, H. (2019). Searching for MobileNetV3. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 1314-1324.
https://doi.org/10.1109/ICCV.2019.00140

[6] Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C.R., Ye, C., Akin, B., Aggarwal, V., Zhu, T., Moro, D., & Howard, A. (2024). MobileNetV4 - Universal Models for the Mobile Ecosystem. European Conference on Computer Vision.

[7] Sharan, R. V., & Moir, T. J. (2019). Acoustic event recognition using cochleagram image and convolutional neural networks. Applied Acoustics, 148, 62-66.
https://doi.org/10.1016/j.apacoust.2018.12.006

[8] Sharan, R. V., Xiong, H., & Berkovsky, S. (2021). Benchmarking Audio Signal Representation Techniques for Classification with Convolutional Neural Networks. Sensors, 21(10), 3434.
https://doi.org/10.3390/s21103434
KDD International Conference on Knowledge Discovery & Data Mining, 2623–2631..

Chau-Chung Song



Chau-Chung Song is with the department of Aeronautical Engineering, National Formosa University, Taiwan. He received the M.S. degree in Control Engineering in 1994 and the Ph.D. degree in Electrical and Control Engineering from National Chiao Tung University in 2001 in Taiwan. He has been with the Department of Aeronautical Engineering, Institute of Aviation and Electronic Technology, National Formosa University, Huwei, Taiwan, since 2006, where he is currently a distinguished professor. His research interests include the dynamic analysis and control of flight systems, the UAV technology and intelligent applications, the cloud networking and information systems, the Cyber-Physical integration and Digital-Twin systems, and the wireless sensor network systems.

Li-Hao Chen

Li-Hao Chen received his Bachelor's degree from the Department of Aeronautical Engineering, National Formosa University, Taiwan in 2022. He is currently a Master’s course student in National Formosa University, Taiwan.

Ping-Chia Wu

Ping-Chia Wu received his Bachelor's degree from the Department of Aeronautical Engineering, National Formosa University, Taiwan in 2025. He is currently a Master’s course student in National Formosa University, Taiwan.